



**You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice**

Title: Gramatyka w komputerze : o dydaktycznych pożytkach płynących ze spotkań lingwistyki z informatyką

Author: Tomasz Nowak

Citation style: Nowak Tomasz. (2016). Gramatyka w komputerze : o dydaktycznych pożytkach płynących ze spotkań lingwistyki z informatyką. W: E. Jaskółowa, D. Krzyżyk, B. Niesporek-Szamburska, M. Wójcik-Dudek; przy współpracy D. Jagodzińskiej i A. Zok-Smoły (red.), "Edukacja polonistyczna jako zobowiązanie : powszechność i elitarność polonistyki. T. 1" (S. 429-445). Katowice : Wydawnictwo Uniwersytetu Śląskiego



Uznanie autorstwa - Użycie niekomercyjne - Bez utworów zależnych Polska - Licencja ta zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu jedynie w celach niekomercyjnych oraz pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Tomasz Nowak

Uniwersytet Śląski w Katowicach

Gramatyka w komputerze O dydaktycznych pożytkach płynących ze spotkań lingwistyki z informatyką

Wstęp

Niniejszy artykuł został pomyślany, a następnie wygłoszony, jako referat w ramach Kongresu Dydaktyki Polonistycznej. Perspektywa, jaką przyjąłem, stanowi punkt widzenia kogoś, kto uczy (od kilkunastu lat) zarówno studentów wyższych uczelni, jak i uczniów szkół podstawowych oraz gimnazjalnych. Przede wszystkim jednak zabieram głos jako ktoś, komu nieobcy jest formalny opis języka, jaki przeprowadza się na gruncie lingwistyki (matematycznej/informatycznej). W tej pracy zamierzam wskazać szereg korzyści, jakie może przynieść uwzględnianie w programie kształcenia i/lub w trakcie lekcji związków łączących gramatykę z matematyką oraz informatyką. Swoje uwagi kieruję do wszystkich osób (zwłaszcza do nauczycieli), którym leży na sercu dobro edukacji polonistycznej, w tym wysoki (wyższy) poziom nauczania języka polskiego (i w szkole, i na uczelni). W swojej pracy ograniczam się jedynie do sformułowania pewnej idei, do poddania pod dyskusję określonego pomysłu — być może „na przyszłość”, być może już „na teraz”.

Sytuacja gramatyki w dydaktyce

W mojej opinii sytuacja gramatyki w dydaktyce przedstawia się źle. Dyscyplinę tę (zarówno w dydaktyce szkolnej, jak i akademickiej) systematycznie poddaje się albo redukcji (w szkole podstawowej i gimnazjalnej), albo eliminacji (w szkole średniej); nie lepiej też sytuacja przedstawia się na wyższych uczelniach. Ze względu na dominację pewnej intelektualnej mody (mam na myśli fiksję uczonych na punkcie tekstu/dyskursu) zaniedbywane staje się kształcenie w zakresie przedmiotów klasycznych (systemowych), stanowiących niezbywalny językoznawczy kanon, w tym także — podstawę formułowania sądów o użyciu języka.

Moja ocena sytuacji gramatyki w dydaktyce jest — i nie zamierzam tego ukrywać — negatywna. Jej względna nieobecność w programach kształcenia jest nieproporcjonalna do odgrywanej przez nią roli i przypisywanej jej rangi. Uważam, i to postaram się dalej wykazać, że gramatyka jest nie tylko dyscypliną najstarszą, ale również źródłową; nie tylko wysoce uteoretyzowaną, ale także praktyczną — o niebywałej skali zastosowań (co przemawia, w moim odczuciu, na rzecz jej większego uwzględniania w „dydaktyce” — na wszystkich szczeblach nauczania).

Gramatyka, o czym się często zapomina (albo po prostu się nie wie), jest najstarszą dyscypliną nauki: zarówno w obrębie językoznawstwa, jak i w uniwersum wszystkich innych gałęzi wiedzy w ogóle (z filozofią włącznie). Przypomnę, że Pāṇini, gramatyk indyjski, ogłosił pierwszy traktat gramatyczny (*Ośmioksiąg*) w V wieku p.n.e., a jego wielcy poprzednicy, np. Yāska, formułowali reguły gramatyczne na długo przed tym, nim traktat ów powstał (co istotne, od samego początku swojego istnienia gramatyka była dziedziną autonomiczną — nigdy nie musiała się wyodrębnić z łona filozofii lub filologii). Na marginesie: poziom, jaki reprezentowała gramatyka staroindyjska, został osiągnięty przez gramatyków europejskich (wzorujących się na filologach rzymskich, zob. Priscianus z Caesarei) dopiero w XIX wieku (po odkryciu dzieł gramatyków indyjskich)¹.

Gramatyka, co należy podkreślić, opisuje (i próbuje wyjaśnić) to, co bezdyskusyjnie wyróżnia człowieka ze świata przyrody ożywionej, mianowicie język jako kod dwuklasowy fonematyczny. Język gromadzi (w słowniku) niewielką liczbę jednostek, które dzięki

¹ A. HEINZ: *Dzieje językoznawstwa w zarysie*. Warszawa 1978, s. 19–26.

operacji proporcji i rekursji można kombinować w przeliczalnie nieskończenie wiele ciągów w taki sposób, że nazwy (znaki pojedyncze) odpowiadają (izomorficznie mapują) przedmiotom, z kolei zdania (znaki złożone) — sytuacjom. Możliwość tę gwarantuje naprzemienna łączliwość członów proporcji: $ab/cd = ad/cb$, a także wynikająca z niej zdolność do zagnieżdżania jednych konstrukcji w drugih, por. funktory $\{a\ c\} _ i _ \{b\ d\}$ oraz produkcje $acac$ i $bdbd$. W takim ujęciu ewolucja od zwierzęcego nie-języka do ludzkiego języka wiąże się z katastroficznym przejściem od reprezentacji pierwszego stopnia w nie-języku do reprezentacji drugiego stopnia w języku. Od tego właśnie momentu na osi czasu datuje się narodziny człowieka².

Gramatyka jest dyscypliną praktyczną, co oznacza, że ma swoje wielorakie zastosowania. W swojej pracy pragnę zwrócić uwagę na niezauważane dotąd w dydaktyce korzyści, jakie płyną z mariażu dwóch dyscyplin, mianowicie: gramatyki lingwistycznej i informatyki matematycznej. Innymi słowy: mam na myśli to, iż model gramatyczny zaimplementowany w pamięci komputera pod postacią programu komputerowego umożliwia ekstrakcję informacji z tekstów, por. rolę, jaką odgrywają obecnie w naszym życiu cyfrowe wyszukiwarki. W tym miejscu nasuwa się pytanie: czyż można sobie współcześnie wyobrazić funkcjonowanie człowieka w informacyjnym szumie (zlokalizowanym w medium języka) bez jakiegokolwiek elektronicznego kompasu czy busoli?.

Człowiek w świecie tekstów i maszyn

Człowiek współczesny żyje w świecie tekstów oraz maszyn. Refleksja nad jednymi i drugimi umożliwia nam wgląd w nas samych. Co więcej, teksty poddają się obróbce przy użyciu maszyn, które umożliwiają penetrację tekstów i ekstrakcję treści. Ten aspekt uważam w swej wypowiedzi za kluczowy (stanowi on wątek przewodni i istotne tło formułowanych tutaj tez).

Zdolności poznawcze pojedynczego człowieka są w gruncie rzeczy znikome: ilość dostępnych w sieci tekstów (i zawartych w nich informacji) znacznie wykracza poza możliwości jednostki, w związku z czym obszerne połacie tekstów pozostają poza jej bezpośrednim

² I. Kurcz: *Psychologia języka i komunikacji*. Warszawa 2000, s. 29—45.

zasięgiem. Co więcej, ilość/jakość wyników, jakie generuje komputer po wpisaniu przez użytkownika do wyszukiwarki poszukiwanej frazy, przytłacza i zniechęca czy wręcz, paradoksalnie, oddala go od poszukiwanej wiedzy. W tym kontekście wyłania się paląca potrzeba skonstruowania poręcznego instrumentu – lingwistyczno-informatycznego „nawigatora”.

Komputer jest narzędziem do przetwarzania informacji (m.in. danych zakodowanych w języku naturalnym): z jednej strony *maszyną matematyczną*, a z drugiej – *mózgiem elektronowym*, co oznacza z grubsza tyle, że pod względem matematycznym imituje sposób funkcjonowania ludzkiego umysłu, por. komputacjonistyczną metaforę komputerową: *komputer/program = mózg/umysł*, tj. analogię *hardware – software*; w tym ujęciu program komputerowy stanowi nie tylko model, ale również – a może przede wszystkim – protezę ludzkiego umysłu.

Cel wysiłków podejmowanych przez informatyków-lingwistów stanowi system informatyczny wyszukujący w sieci internetowej informacje potrzebne użytkownikowi oraz przekładający teksty z jednego języka na inne, użytkownikowi nieznane. Realizacja tychże zadań wymaga specjalnego opisu szukanej informacji, zrozumiałego tak dla komputera, jak i dla użytkownika, umożliwiającego mu odnalezienie żądanej informacji w skończonym czasie (pamięć użytkownika) i ograniczonej przestrzeni (pamięć komputera). W tym miejscu spotykają się ze sobą informatyka z lingwistyką. Fuzja informatycznych narzędzi i lingwistycznych teorii (w coraz większym stopniu) umożliwia eksplorację tekstów w sieci bądź ekstrakcję informacji z tekstów (dla języka polskiego istnieje obszerny, dostępny w sieci *online* i *offline*, zbiór instrumentów lingwistyczno-informatycznych, zob. <http://clip.ipipan.waw.pl/LRT>).

Młody człowiek – uczeń lub student – żyje w świecie tekstów, które stwarzają i kształtują jego naturalne środowisko. Młody człowiek – *cyfrowy tubylec* – posługuje się również urządzeniami elektronicznymi, które niemal na każdym kroku (wszak nieprzerwanie znajduje się w sieci: *online*) stanowią naturalne (fantomowe) przedłużenie jego ciała i ducha. Należy się spodziewać, że sytuacja ta nie ulegnie zmianie, por. opisywany szeroko *efekt kulturowej zapadki*. Myślę więc (w kontekście dydaktyczno-polonistycznym), że najwyższa pora stawić czoło wyzwaniom cyfrowej rzeczywistości i po prostu wykorzystać (połączyć) tkwiące w dzieciach i młodzieży kompetencje i sprawności: językowo-komunikacyjne oraz informatyczne.

W kolejnych ustępach zastanowię się nad możliwościami wykorzystania w dydaktyce szkolnej (i uczelnianej) wybranych — odpowiednio spopularyzowanych i podanych — osiągnięć lingwistyki matematycznej oraz informatycznej.

Lingwistyka matematyczna i informatyczna w dydaktyce

Język w równaniach

Lingwistyka matematyczna, charakteryzując elementarne i uniwersalne właściwości języków naturalnych, zakłada, że u ich podstaw tkwią pewne prawidłowości i regularności. Językoznawcy matematyczni podejmują próby formułowania skończonego zbioru reguł lub restrykcji, za pomocą których można konstruować „przeliczalnie nieskończoną ilość przeliczalnie nieskończone długich zdań”. Moc syntaktyczna języka tkwi w regułach, które mają tę własność, że odwołują się same do siebie, dzięki czemu można je realizować (rekurencyjnie) wiele razy. Ponieważ liczba kombinacji składników w zdaniach jest nieograniczona, język, tak jak algebra, otwiera ogromne możliwości. Sytuację tę znakomicie ilustruje następująca obserwacja: otóż 10 wyrazów może wejść ze sobą w 3 628 800 kombinacji, przy czym w zbiorze tym ledwie kilka kombinacji słów przypomina zdania; lingwiści-matematycy stawiają sobie właśnie za cel dotarcie do reguł odróżniających owe zdania od nie-zdań.

Lingwistyka matematyczna rozwija się w dwóch kierunkach: matematycznym (teoria języków formalnych) i lingwistycznym (teoria języków naturalnych). Zadania, jakich podejmują się uczeni (lingwistycznie ukierunkowani matematycy i matematycznie ukierunkowani lingwiści), ogniskują się na konstrukcji języków formalnych (reprezentacji skończonych) dla opisu języków naturalnych (obiektów nieskończonych). Gramatyki i automaty generują (odp.: syntetyzują i analizują) języki formalne, które — w zamysłach badaczy — mogą pełnić funkcję modeli teoretycznych dla języków naturalnych. Co istotne, języki formalne (jak również: generujące je gramatyki oraz rozpoznające je automaty) tworzą, ze względu na cechy organizujących je reguł, hierarchię (od najsilniejszych do najsłabszych): REG < CF < CS < RP, odp. języki: regularne, bezkontekstowe, kontekstowe, rekurencyjnie przeliczalne. Na tej podstawie

udowodniono m.in., że języki naturalne nie są językami regularnymi; strukturę syntaktyczną języków naturalnych zadowalająco modelują gramatyki bezkontekstowe³.

Gramatyka — w ujęciu matematycznym czwórka uporządkowana: Σ, V, δ, P — obejmuje cztery zbiory, mianowicie: zbiór reguł (P) oraz zbiory symboli: podstawowych (alfabet terminalny: Σ) i pomocniczych (alfabet nieterminalny: V), w tym symbol inicjalny (δ). Ze względu na możliwość przyporządkowywania poszczególnym symbolom różnych interpretacji, gramatyki, które opisują budowę języków naturalnych, rozpadają się na dwie klasy: frazowe (PSG) i dependencyjne (DG). Co istotne, gramatyki: PSG i DG pozostają do siebie w relacji słabej równoważności, a to oznacza, że generują identyczny zbiór zdań za pomocą różnych rozbiorów (przypisując zdaniom odmienne struktury). Czym różnią się te dwie klasy modeli gramatycznych? Czy preferowanie jednej z nich (kosztem drugiej) w nauczaniu jest uprawnione?⁴.

Gramatyka dependencyjna opisuje zależności między słowami — rządzącymi i rządzonymi, tj. modyfikatorami — oparte na relacji predykatywności (diagnozowanej za pomocą testu pytań) jako konstrukcje endocentryczne. Reprezentacja struktury zdania przyjmuje postać drzewa zależności, w którym korzeń-wierzchołek markuje orzeczenie, węzły — słowa, a liście — słowa etykietywane jako funkcje syntaktyczne (części zdania), por. przykładowy opis zdania: *Adam polubił Ewę.*: $N1 \leftarrow \text{subj} (1) \leftarrow \text{pred} \rightarrow \text{obj} (2) \rightarrow N2$.

Gramatyka frazowa charakteryzuje dominację między frazami (nadrzędnymi i podrzędnymi, tj. podrzędnikami), zasadzające się na stosunku podrzędności (identyfikowanym za pomocą testu redukcji), jako konstrukcje endo- oraz egzocentryczne. Reprezentacja struktury zdania przybiera kształt drzewa składników, w którym: korzeń-wierzchołek fiksuje zdanie, węzły — frazy, a liście — słowa zakwalifikowane jako klasy gramatyczne (części mowy), zob. opis przykładowego zdania: *Adam polubił Ewę.*: $S \rightarrow NP + VP$; $NP \rightarrow N$; $VP \rightarrow V + NP$.

Na podstawie najważniejszych założeń obydwu klas gramatyk: frazowej (amerykańskiej) oraz dependencyjnej (europejskiej) skonstruowano kilkaset modeli gramatycznych. W gruncie rzeczy nadal jednak nie wiemy, który z nich wiernie opisuje gramatykę: taką, jaka rzeczywiście rezyduje w naszych mózgach/umysłach (jako moduł

³ A. MYKOWIECKA: *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*. Warszawa 2007, s. 23–63.

⁴ Ibidem, s. 85–125.

lub sieć). „Skandal językoznawstwa” polega więc na tym, że po 2 500 latach badań wciąż nie wiadomo, jaką strukturę mają proste zdania w rodzaju *Adam polubił Ewę*. Problem, jaki tu pragnę zauważyć, najlepiej oddaje pytanie: jakiej gramatyki, w związku z tym wszystkim, uczyć (w szkołach i na uczelniach)?. (Na marginesie: badania, jakie prowadzi się w dziedzinach: psycho- i neurolingwistyki, wskazują raczej na gramatyki frazowe — jako bardziej realne: psychologicznie i neuronalnie).

Gramatyka szkolna (klasyczna i tradycyjna), dominująca w dydaktyce (zarówno szkolnej, jak i akademickiej), nawiązując do grecko-lacińskiej refleksji filologicznej, wpisuje się w nurt modeli dependencyjnych. Wiele krytycznych słów⁵ napisano pod adresem kompletności i eksplicytności dostarczanego przez nią opisu polskiej składni. Nie zamierzam powtarzać szeroko znanych uwag i krytyk. Nie mogę jednak nie zauważyć, że gramatyka szkolna nie rejestruje wielu relewantnych syntaktycznie zjawisk — zwłaszcza opartej na kategorii funktora implikacji. Nie mogę również nie wspomnieć, że gramatyka klasyczna/tradycyjna jest koncepcją: nieadekwatną (nie zdaje sprawy z wielu najistotniejszych mechanizmów rządzących językiem), nierozwojową (nie jest rozwijana w żadnym znanym naukowym ośrodku) i nieimplementowalną (nie jest możliwa jej formalizacja i instalacja w pamięci komputera). Przyczyn tego stanu rzeczy można upatrywać w tym, że teoria ta reprezentuje podejście eklektyczne, w którym łączą się, często w sposób nieuprawniony, treści syntaktyczne, semantyczne i pragmatyczne. Podstawowe zasady gramatyki szkolnej kształtowały się w drugiej połowie XIX wieku. Sytuacja polonisty, który ogranicza się do znajomości tylko tej koncepcji — i tylko jej założenia przekazuje młodym adeptom, przypomina sytuację fizyka, który — opowiadając się za ustaleniami fizyki klasycznej — zaprzeczałby, jakoby w XX wieku w fizyce wydarzyło się coś przełomowego. Dla badań nad językiem ostatnie stulecie było przełomowe do tego stopnia, że uczeni niemal zgodnie okrzyknęli miniony wiek okresem dominacji paradygmatu lingwistycznego w dyscyplinach społeczno-humanistycznych. W związku z tym uważam, że nie można w procesie dydaktycznym nie uwzględniać (pomijać) osiągnięć dwudziestowiecznych, np. w zakresie gramatyki języków naturalnych.

Matematyczna refleksja nad językiem zmierza w dwóch kierunkach: algebraiczno-logicznym (jakościowym) i statystyczno-pro-

⁵ M. ŚWIDZIŃSKI: *Gramatyka formalna języka polskiego*. Warszawa 1992, s. 32–51.

babilistycznym (ilościowym). W swoim szkicu nawiązę do tego dwupodziału, proponując wykorzystanie w dydaktyce gramatyk — kategoryalnej i deszyfracyjnej, które pod wieloma względami stanowią przeciwieństwo naukowo zarzuconej gramatyki szkolnej. Teorie syntaktyczne, jakie sugeruję wprowadzić do obiegu dydaktycznego, eksponują ważne (być może najważniejsze) rysy budowy języków naturalnych: funktorowość (gramatyka kategoryalna) i przewidywalność (gramatyka deszyfracyjna).

Gramatyka kategoryalna — mądrzej głowie dość dwie słowie

Gramatyka kategoryalna w swojej najstarszej i najprostszej wersji⁶ ogniskuje się wokół dwóch zjawisk syntaktycznych: kategorii i algorytmu. Przybliżę je w kolejnych ustępach.

Kategoria syntaktyczna to, ogólniej rzecz ujmując, klasa wyrażen subystytuujących. W zbiorze kategorii syntaktycznych można wyodrębnić kategorie podstawowe, a więc nazwy i zdania, oraz niepodstawowe, czyli funktory, przy czym: nazwa jest wyrażeniem, które nadaje się na podmiot lub orzecznik, zdanie jest wyrażeniem, któremu można przypisać wartość logiczną, a funktor jest wyrażeniem odgrywającym rolę funkcji, która wyrażeniom prostym i nieproszym (argumentom) przyporządkowuje wyrażenia złożone (wartości). Co istotne, kategorie podstawowe notuje się za pomocą wskaźnika pojedynczego (np. *n*, *s*), a kategorie niepodstawowe — przy użyciu wskaźnika ułamkowego (np. *n/s*), w którym mianownik wyklada argument, a licznik — wartość konstrukcji, por. przykład: (1) pozytywny: *Adam polubił Ewę.*, gdzie: kategoria: *Adam: n*, *polubił: s/nn*, *Ewę: n*; (2) negatywny: *Adam polubił.*, gdzie: kategoria: *Adam: n*, *polubił: s/nn*.

Algorytm syntaktyczny umożliwia weryfikację poprawności (spójności) składniowej zdań: po pierwsze, przyporządkowuje wyrażeniom języka symbole kategorii, po drugie, zapisuje symbole kategorii (zgodnie z szykiem wyrażen w zdaniu) i po trzecie, redukuje powtarzające się w zdaniu symbole kategorii (równokształtne, mianownikowe, sąsiadujące). Finalnie, algorytm rozpoznaje zdanie jako zbudowane poprawnie, pod warunkiem, że ciąg symboli kategoryalnych redukuje się do wskaźników: pojedynczego (kategoria podstawowa) lub ułamkowego (kategoria funktorowa), por. przykład: (1) pozytywny: *Adam polubił Ewę.*, gdzie algorytm: ciąg:

⁶ Na podstawie: K. AJDUKIEWICZ: *Język i poznanie*. T. 1. Warszawa 1985, s. 222–242.

n , s/n , n ; sekwencja: $n + s/nn + n$; redukcja: $n + s/nn + n \rightarrow s/n + n \rightarrow s$; wynik: wyrażenie spójne syntaktycznie; (2) negatywny: *Adam polubił.*, gdzie algorytm: ciąg: n , s/nn , sekwencja: $n + s/nn$, redukcja: $n + s/nn \rightarrow s/n$, wynik: wyrażenie niespójne syntaktycznie.

Podsumowując: gramatyka kategoryalna jest koncepcją syntaktyczną: prostą — odwołuje się do zaledwie dwóch kategorii, jednorodną — operuje kryteriami wyłącznie składniowymi, rodzimą — powstała i rozwija się w Polsce, poddającą się aksjomatyzacji, akceptującą relewantne językowo zasady: funktorowości i kompozycyjności.

Gramatyka deszyfracyjna — „stwórz własną gramatykę”

Gramatyka deszyfracyjna w swej najprostszej i najbardziej popularnej wersji⁷ przewiduje dwa etapy docierania do struktury składniowej zdań: deszyfrację gramatyki (z tekstu) i analizę zdania (przy użyciu wyekscerpowanej z tekstu gramatyki). Omówię je w kolejnych akapitach.

Deszyfracja syntaktyczna to operacja budowy gramatyki, przebiegająca w dwóch krokach. W kroku pierwszym wybiera się tekst wielozdaniowy, który umożliwi rekonstrukcję rządzących nim zasad gramatycznych, np. *Adam poznał niedawno jasnowłosą Ewę. Sympatyczny chłopiec gra w piłkę. Miła dziewczynka śpiewa piosenki. Adam polubił Ewę. Teraz dzieci grają na pianinie* itd. W kroku drugim analizowany tekst zostaje opisany ogólnie w taki sposób, że przyporządkowuje się każdej formie symbol klasy, do której forma przynależy (tj. etykietę części mowy), np.: *Adam_N poznał_V niedawno_{ADV} jasnowłosą_{ADJ} Ewę_N. Sympatyczny_{ADJ} chłopiec_N gra_V w_P piłkę_N. Miła_{ADJ} dziewczynka_N śpiewa_V piosenki_N. Adam_N polubił_V Ewę_N. Teraz_{ADV} dzieci_N grają_V na_P pianinie_N*. W kroku trzecim buduje się tabelę, której wiersze i kolumny etykietują symbole klas form, użyte w opisie tekstu. Następnie kolejne pary (symboli klas) form rozgrywają ze sobą pojedynki. Rzecz jasna, niektóre pary są akceptowalne (wygrana), a niektóre (co sygnalizuje potępiająca gwiazdka) — nieakceptowalne (przegrana), np. *Adam poznał*, **Adam niedawno*, **Adam jasnowłosą*, **Adam Ewę*, *poznał niedawno*, **poznał jasnowłosą*, *poznał Ewę*, **niedawno jasnowłosą*, **niedawno Ewę*, *jasnowłosą Ewę*. Ostatecznie zlicza się rezultaty na zasadzie: para zwycięska *vs* para przegrana. Wyniki rozgrywek wpisuje się — jako liczby (wygrana = 1 lub przegrana = 0) — do

⁷ Na podstawie: J. APRESJAN: *Koncepcje i metody współczesnej lingwistyki strukturalnej*. Przeł. Z. SALONI. Warszawa 1971, s. 159—195.

rubryk tabeli, a następnie sumuje. Otrzymuje się w rezultacie liczby: większe i mniejsze. Wartość liczb (większa lub mniejsza) oddaje siłę związku gramatycznego (większą lub mniejszą) łączącego kolejne pary form. W ten sposób koduje się w tabeli gramatykę probabilistyczną, czyli informacje (o relacjach syntaktycznych w zdaniu) ilościowe przekładalne na jakościowe (zob. tabela 1.).

Tabela 1

Gramatyka probabilistyczna przykładowego tekstu

	Przymiotnik ADJ	Rzeczownik N	Czasownik V	Przyimek P	Przysłówek ADV
Przymiotnik ADJ	0	3	0	0	0
Rzeczownik N	0	0	5	0	0
Czasownik V	0	3	0	2	1
Przyimek P	0	2	0	0	0
Przysłówek ADV	0	0	1	0	0

Źródło: Opracowanie własne.

Analiza syntaktyczna to etap aplikacji probabilistycznej gramatyki do konkretnego przykładu. W kroku pierwszym sporządza się graf spójny zdania, łącząc ze sobą za pośrednictwem krawędzi wszystkie formy. W kroku drugim wpisuje się nad krawędziami grafu odpowiednie liczby, wzięte z rubryk tabeli. W kroku trzecim eliminuje się kolejne krawędzie grafu: zaczynając od tych, nad którymi jest zapisana liczba najmniejsza. W ten sposób otrzymuje się drzewo ilustrujące strukturę syntaktyczną analizowanego zdania, por.: krok₁: graf: *Adam* —1 *polubił* —2 *Ewę* —3 *Adam*; krok₂: krawędź₁: *Adam polubił* (5); krawędź₂: *polubił Ewę* (3); krawędź₃: *Adam Ewę* (0); krok₃: *Adam* (5) → *polubił* (3) → *Ewę*.

Postępując konsekwentnie, uczeń/student może skonstruować gramatykę, kierując się jedynie umiejętnością zliczania rezultatów rozgrywek między parami form w zdaniach wziętych z tekstu. Uczeń, (re)konstruując gramatykę zawartą „między wierszami” w tekście, poznaje podstawowe zasady rządzące strukturą syntaktyczną (każdego) języka.

Język w komputerze

Kwestie związane z przetwarzaniem komputerowym języka naturalnego podejmuje się przede wszystkim na gruncie dwóch dziedzin: lingwistyki informatycznej (komputerowej) i informatyki (inżynierii) lingwistycznej. Obie dyscypliny różnią się od siebie pod względem stawianych celów i stosowanych metod. O ile bowiem lingwistyka informatyczna podejmuje cele teoretyczne (m.in. większą wagę przywiązuje do adekwatności empirycznej — precyzji i prawdziwości opisu), o tyle informatyka lingwistyczna realizuje zadania praktyczne (większą uwagę zwraca na efektywność obliczeniową — szybkość działań i praktykę zastosowań). Innymi słowy: lingwista informatyczny konstruuje programy, które — na podstawie informacji o języku, zapisanych pod postacią reguł w metajęzyku sformalizowanej teorii lingwistycznej — symulują i emulują funkcjonowanie systemu językowego; informatyk lingwistyczny tworzy natomiast aplikacje umożliwiające: wyszukiwanie informacji w sieci, tłumaczenie automatyczne oraz syntezę/analizę mowy (programy, jakie pisze, przetwarzają teksty w poszukiwaniu kodowanych w językach naturalnych informacji o świecie). Co więcej, podczas gdy lingwistyka informatyczna rozwija się w granicach nurtu symbolicznego-formalnego (generując język z pomocą gramatyk bezkontekstowych), informatyka lingwistyczna postępuje w ramach nurtu stochastycznego-statystycznego (wykorzystując do tego celu automaty niedeterministyczne). Współcześnie zaznacza się silna tendencja do łączenia obu paradygmatów. Najogólniej rzecz ujmując, cel, jaki stawiają sobie lingwiści-informatycy, sprowadza się do opisu (jakiegoś aspektu) języka naturalnego za pomocą (któregoś z wielu) języka programowania (tak na marginesie: nie mogę się oprzeć wrażeniu, że pod niektórymi względami językoznawcy oddali pole informatykom; tak się bowiem składa, że najbardziej zaawansowane teorie budowy języka są obecnie formułowane — i implementowane — w środowisku inżynierów, nie — humanistów)⁸.

Co z tego wynika dla nauczyciela oraz ucznia/studenta? Jak objaśniać młodemu człowiekowi złożone relacje, jakie zachodzą między używanymi przez niego — biegle i równolegle — językami: naturalnym i komputerowym (wszak, o czym nauczyciele zapominają, większość młodych ludzi potrafi — w mniejszym bądź więk-

⁸ Więcej na ten temat w: M. PIASECKI: *Cele i zadania lingwistyki informatycznej*. W: *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje*. Red. P. STAŁMASZCZYK. Kraków 2008, s. 252—290.

szym stopniu — programować urządzenia elektroniczne)? Myślę, że warto wyjść od porównania ze sobą obu kodów (odwołując się przy tym do technik przekładu intersemiotycznego).

Pomiędzy językami ludzi i komputerów istnieje (pod wieloma względami) mnóstwo analogii, które warto uczniowi (w szerszej semiotyczno-semiologicznej perspektywie) uzmysłowić. Myślę, że dzięki temu łatwiej pojmie (czy wręcz sobie unaoczní) zarówno naturę języka, jak i wszelkich wytworów cywilizacji, które noszą w sobie niezbywalne językowe piętno.

Po pierwsze, języki komputerów operują jednostkami, dla których można bez większego trudu wskazać analogie w językach ludzi, por. język komputerowy: *bit, bajt, słowo, rekord, plik, folder* i ludzki: *litera, sylaba, wyraz, zdanie, tekst, biblioteka*. Po drugie, języki komputerów, dokładnie tak jak języki ludzi, obejmują słownik i gramatykę, por. język komputerowy: *alfabet (symbole) i instrukcje (formuły)* oraz ludzki: *leksyka (wyrazy) i reguły (zdanía)*. Po trzecie wreszcie, języki: komputerów i ludzi mają (ściśle określoną) budowę, o której precyzję dbają komputerowe: *translatory symboli i kompilatory formuł*, jak również ludzkie: *reguły gramatyczne* (na podstawie części mowy w słowniku) i *funkcje logiczne* (w oparciu o części zdań w wypowiedzeniach).

Języki: ludzkie i komputerowe łączy sporo podobieństw oraz dzieli sporo różnic. Po pierwsze, ludzie rozmawiają (ze sobą) w językach naturalnych: „wewnętrznych”, por. myślenie (w formie monologu — mowa do siebie), oraz „zewnętrznych”, por. mówienie (w formie dialogu — mowa do innych). Po drugie, komputery rozmawiają ze sobą w językach programowania „wewnętrznych” (języki niskiego poziomu), jak też z ludźmi — w językach programowania „zewnętrznych” (języki wysokiego poziomu), przypominających swoją strukturą języki naturalne, por. języki zewnętrzne (dedykowane wielu komputerom) wysokiego poziomu, np. język europejski Algol (w tym Pascal i Modula), a także język amerykański Fortran (w tym C i Logo)⁹. Rzecz jasna, komputer to jedynie poręczne narzędzie penetracji tekstów i selekcji informacji: sposób orientacji w świecie tekstów. Nie można więc zapominać, że te fascynujące maszyny to tylko użyteczne instrumenty, nie bałwochwalcze fetysze.

Program nauczania informatyki w gimnazjum i liceum przewidyuje ćwiczenia w zakresie prób algorytmizacji i programowania.

⁹ A. FAUDROWICZ: *Komputery i my. Zarys informatyki dla laików*. Warszawa 2001.

Nabyte w trakcie zajęć informatycznych umiejętności można by łączyć z wiedzą na temat gramatyki, przyswajaną podczas lekcji polonistycznych. Przypomnę, że algorytm (termin zadomowiony w dydaktyce polonistycznej) to zbiór operacji, które umożliwiają rozwiązanie problemu (tj. przejście od „danych” do „szukanych” dzięki odpowiednim operacjom i reprezentacjom), imitujący sposób myślenia człowieka. Istnieje szereg publikacji zawierających przykładowe przepisy algorytmy dla gramatyki języka polskiego, przeznaczone do użytku przez nauczycieli i uczniów, a nawet studentów¹⁰, zob. np. algorytm rozpoznający bezokoliczniki:

Operacja 1: Czytaj pierwszy element tekstu.

Operacja 2: Sprawdź, czy element ten kończy się na „ć”.

Jeżeli tak, to przejdź do operacji 3.

Jeżeli nie, to przejdź do operacji 4.

Operacja 3: Sprawdź, czy element jest na liście wyjątków.

Jeżeli tak, to przejdź do operacji 6.

Jeżeli nie, to dopisz do sprawdzanego elementu wyraz TAK i przejdź do operacji 6.

Operacja 4: Sprawdź, czy element kończy się na „c”.

Jeżeli tak, to przejdź do operacji 5.

Jeżeli nie, to przejdź do operacji 6.

Operacja 5: Sprawdź, czy element jest na liście wyjątków.

Jeżeli nie, to dopisz do sprawdzanego elementu wyraz NIE i przejdź do operacji 6.

Jeżeli tak, to dopisz do sprawdzanego elementu wyraz TAK i przejdź do operacji 6.

Operacja 6: Usuń z tekstu pierwszy element; przejdź do operacji 7.

Operacja 7: Sprawdź, czy tekst zawiera przynajmniej jeden element.

Jeżeli tak, przejdź do operacji 1.

Jeżeli nie, zakończ pracę.

Algorytm przetłumaczony na określony język programowania to program komputerowy (ściślej: cyfrowa implementacja algorytmu, przyjmująca postać instrukcji: warunkowych i sterujących — w rodzaju sekwencji, pętli, warunku, skoku), por. program, napisany w języku LISP, realizujący algorytm rozpoznający bezokoliczniki¹¹:

¹⁰ T. NOWAK: *Algorytmizacja gramatyki i ortografii — teoria oraz jej wykorzystanie w praktyce*. „Roczniki Humanistyczne KUL” 2004, t. 52, z. 6, s. 149—180; IDEM: *Powtórka z gramatyki. Gimnazjum*. Warszawa 2004; IDEM: *Powtórka z gramatyki. Szkoła podstawowa*. Warszawa 2004.

¹¹ W. BYRSKI, W. LUBASZEWSKI: *LISP — podręcznik programowania dla humanistów*. Lublin 1987, s. 8—18.


```
(DEF
(BADAJ/TEKST)
(MAPCAR TEKST/QUOTE BEZOKOLICZNIK)))
(DEF
(BEZOKOLICZNIK/WYRAZ)
(COND//AND/CI*KONCOWKA WYRAZ/NOT/CI*WYJATEK
WYRAZ)))
/LIST WYRAZ/QUOTE TAK)))
((AND/CE*KONCOWKA WYRAZ/CE*WYJATEK WYRAZ))
(LIST WYRAZ/QUOTE TAK)))
(T/LIST WYRAZ/QUOTE NIE))))))
```

Lingwistyka informatyczna i informatyka lingwistyczna dostarczają elektronicznych urządzeń, a ściślej: zasobów i narzędzi, które mogą się okazać pomocne w procesie dydaktycznym. Część z nich jest powszechnie dostępna, np. w zgromadzonych w Internecie zbiorach IPI PAN. Twierdzę, że uczniów/studentów może zainteresować fakt, iż komputer potrafi — w wielu wypadkach nawet sprawniej niż człowiek — rozwiązywać gramatyczne i leksykalne łamigłówki. Być może wiedzeni ciekawością zechcą samodzielnie przetestować cyfrowe narzędzia i zasoby, zadając komputerowi coraz to trudniejsze przykłady (w ten sposób, obcując z wynikami jego analiz, tj. bawiąc się „(z) komputerem i gramatyką”, zbliżą się do wiedzy o języku). Myślę, że warto na koniec napomknąć kilka słów o zasobach i narzędziach informatyczno-lingwistycznych, po które mogliby sięgać tak nauczyciele, jak i uczniowie¹².

Zasoby to pasywne zbiory danych, które opisują język w różnych jego aspektach, np. słowniki elektroniczne i korpusy tekstowe. Słowniki elektroniczne utrwalają gramatyczne i semantyczne atrybuty leksemów, w tym ich wymogi: semantyczne i morfosyntaktyczne. Korpusy językowe to z kolei zbiory tekstów (wraz z przypisanymi im indeksami), które umożliwiają przeszukiwanie i gromadzenie informacji (pod wieloma względami — w zależności od stawianych celów).

Narzędzia to aktywne zbiory instrukcji, które implementują język na różnych jego poziomach, np. analizatory morfologiczne (tagery) i analizatory syntaktyczne (parsery). Omówię je po kolei.

Analizator morfologiczny (tager) jest programem przeprowadzającym analizę morfologiczną: na podstawie informacji o klasach gramatycznych (tagset) generuje informacje morfologiczne na temat

¹² Na podstawie: M. PIASECKI: *Cele i zadania lingwistyki informatycznej...*, s. 252–290.

segmentów. Analizator wykonuje swoje zadania w trzech etapach: po pierwsze, segmentuje tekst (tokenizacja), po drugie, analizuje segmenty tekstowe (deskrypcja), po trzecie, selekcjonuje interpretacje (dezambiguacja). Rezultat pracy analizatora stanowią — dla wejściowego segmentu — jego forma kanoniczna (lematyzacja), forma bazowa (*stemming*), klasa gramatyczna i kategorie morfosyntaktyczne (klasyfikacja), a także ich wartości, współrzędne paradygmatyczne (anotacja).

Analizator syntaktyczny (parser) jest programem przeprowadzającym analizę syntaktyczną: na podstawie informacji morfologicznych o segmentach generuje informacje syntaktyczne na temat struktury ciągów segmentów. Analizator pracuje w trzech etapach. Jak wspomniałem, wejście do analizatora stanowi opis morfologiczny segmentów, natomiast wyjście — opis syntaktyczny zdań (najczęściej w postaci drzewa), a wreszcie tym, co poziomy te łączy, są zaimplementowane w jego pamięci (stworzone przez lingwistów ręcznie lub wydobyte z tekstów automatycznie) reguły i/lub restrykcje językowe. Analizator pracuje w trybach: płytkim i głębokim, dostarczając opisów zdań ogólnych lub szczegółowych. Opis, jaki analizator proponuje, zależy — rzecz jasna — od modelu gramatycznego, jaki został zaimplementowany jako część programu w pamięci komputera.

Podaję wyniki pracy analizatorów: morfologicznych i syntaktycznych dla segmentów z przykładowego (najprostszego) zdania *Adam polubił Ewę* (obrazy: 1., 2.).

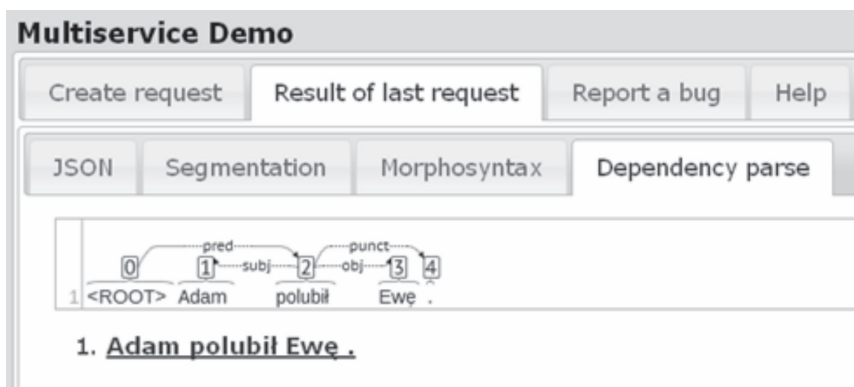
Analizator morfologiczny *Morfeusz SGJP*

Podaj tekst:

0	1	Adam	Adam	subst:sg:nom:m1	imię
1	2	polubił	polubić	praet:sg:m1.m2.m3:perf	
2	3	Ewę	Ewa	subst:sg:acc:f	imię
3	4	.	.	interp	

Obraz 1. *Morfeusz SGJP* (autorzy: Włodzimierz Gruszczyński, Zygmunt Saloni, Marcin Woliński, Robert Wołosz), wersja demonstracyjna

Źródło: IPI PAN, <http://sgjp.pl/morfeusz/demo/> [data dostępu: 5.09.2016].



Obraz 2. *Polish Dependency Parser* (autorka: Alina Wróblewska)

Źródło: IPI PAN, <http://zil.ipipan.waw.pl/PolishDependencyParser> [data dostępu: 5.09.2016].

Przykłady pracy analizatorów morfosyntaktycznych opieram na programach: *Morfeusz SGJP* (Gruszczyński, Saloni, Woliński, Wołosz) i *Polish Dependency Parser* (Wróblewska) — powszechnie dostępnych pod podanym niżej adresem. **Dla języka polskiego istnieje obszerny, dostępny w sieci online i/lub offline zbiór instrumentów lingwistyczno-informatycznych, zob. <http://clip.ipipan.waw.pl/LRT> [data dostępu: 5.09.2016].**

Zakończenie

Relacje między lingwistyką, matematyką oraz informatyką rysują się następująco: lingwistyka dostarcza teorii językowej (np. PSG, DG), matematyka — systemu formalnego (np. REG, CF, CS, RP), a informatyka — programu komputerowego (np. LISP, PROLOG, PERL). W związku z tym nasuwają się trzy pytania: po pierwsze, jaka teoria najlepiej opisuje język?; po drugie, jaki system umożliwia wyczerpującą formalizację tej teorii?; po trzecie, jaki program skutecznie implementuje ten opis i tę formalizację?. Na te pytania — pod pewnymi względami fundamentalne (nie tylko dla refleksji nad językiem) — nie znaleziono, jak na razie, zadowalającej odpowiedzi. Mam nadzieję, że młodzi adepci (uczniowie i studenci), zaznajomieni z rudymetami gramatyki i informatyki, rozwiążą w przyszłości (nie tylko) ten problem. Przed nauczycielami rysuje się problem zgoła inny: jak nadażyć za informatycznymi kompetencjami/sprawnoś-

ciami dzieci i młodzieży?; w jaki sposób połączyć „stare” i „nowe” — z poszanowaniem dla tradycji i w trosce o przyszłość?.

Zagadnienia, jakie poruszyłem, wymagają większego namysłu. Ze swojej strony podzieliłem się jedynie pewną ideą, której trzeba nadać bardziej realny, zinstytucjonalizowany kształt. Jestem przekonany co do tego, że z biegiem czasu lingwistyczno-informatyczne wątki będą coraz silniej infiltrować programy kształcenia. Rozważenie zasadności lub bezzasadności sformułowanych w tej pracy dydaktycznych propozycji pozostawiam życzliwemu Czytelnikowi.

Tomasz Nowak

Grammar on the computer
The educational benefits resulting from meetings of linguistics
with computer science

Summary

The article presents the current situation of grammar in teaching. The author discusses the role that mathematical and computational linguistics should, in his opinion, play in modern school (among others in the Polish language lessons). His thesis is illustrated by numerous examples of grammatical models and computer programs.

Keywords: teaching, grammar, mathematics, computer science

Томаш Новак

Грамматика в компьютере
О дидактической пользе встреч лингвистики
с информатикой

Резюме

В статье показана актуальная ситуация грамматики в дидактике. Автор рассматривает роль, какую математическая и информатическая лингвистика должна, по его мнению, играть в современной школе (в частности на уроках польского языка и литературы). Свои тезисы автор иллюстрирует многочисленными примерами грамматических моделей и компьютерных программ.

Ключевые слова: дидактика, грамматика, математика, информатика